Pet Food Sentiment Analysis

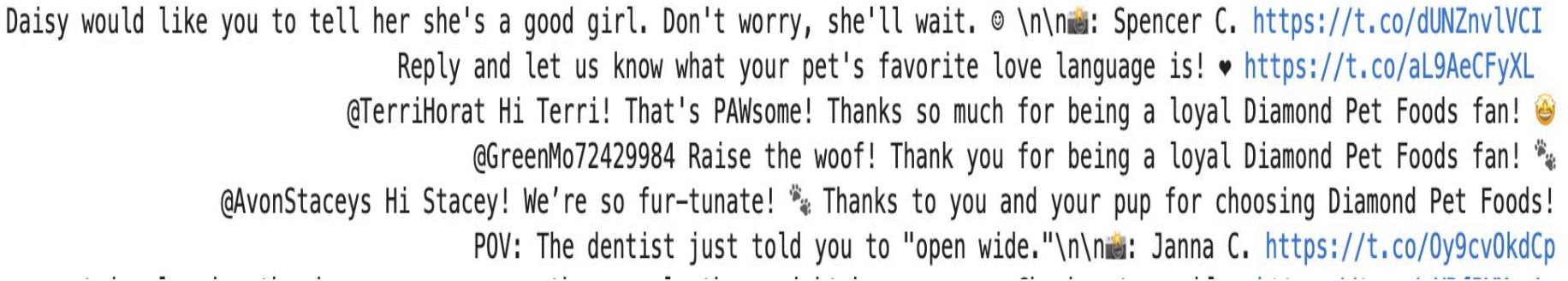
A look at Pet Food Company Tweets
by Kirk Hazen, PhD

Goals

- Assess tweets by four major pet food companies
- Focus on features that might correlate with increased positive sentiment
- Explore linguistic qualities with NLP for rhetorical outreach
- Test out ML models that might predict positive sentiment apart from words themselves

Data Wrangling

Twitter Data*



*Data drawn from Twitter Feb 13th before changes to the Twitter API

Text Crafting

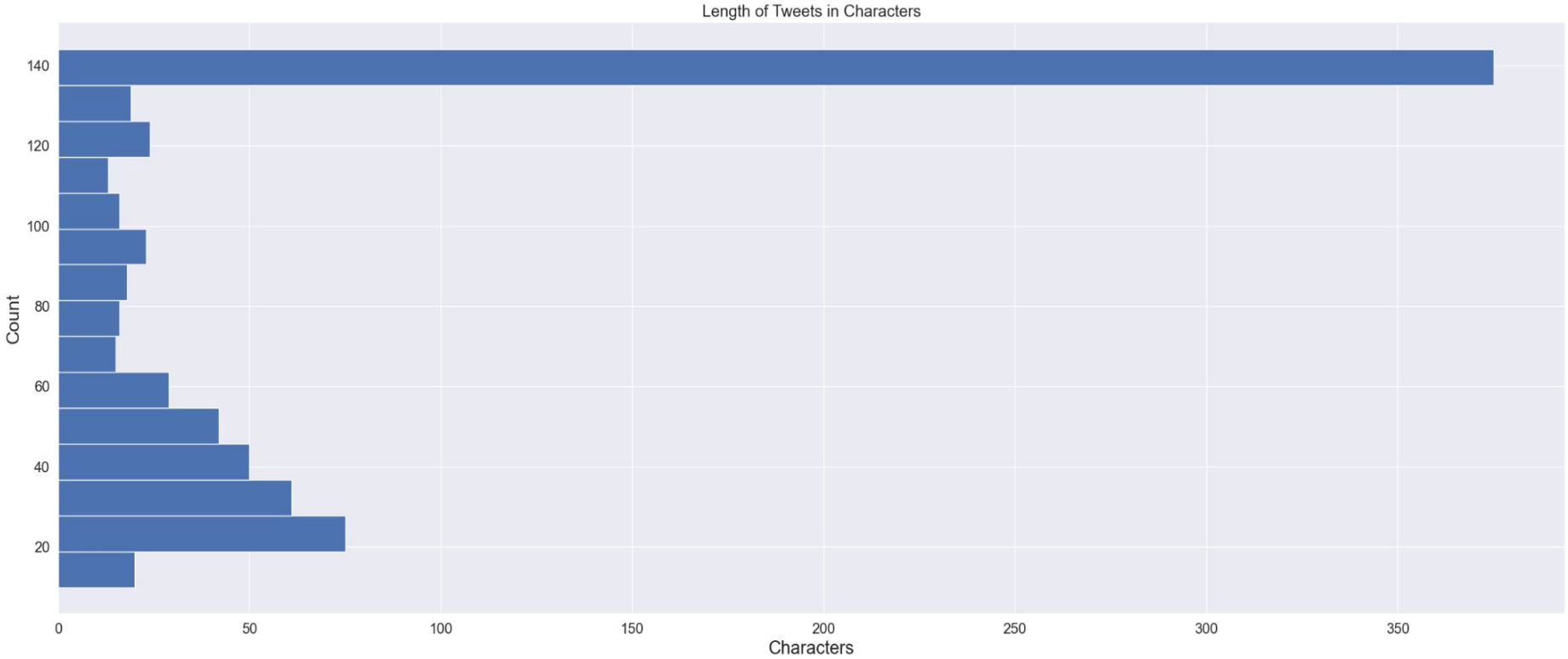
- Images were not captured as part of this NLP assessment
- URLs & Emojis were removed
- Mentions were counted for each tweet as a feature
    - But mentions themselves were filtered out of the full text tweets
- User_Name was created for each company: Blue Buffalo, Diamond Pet Food, Purina, Wellness Pet Food
- Tweets were lowercased
- Contractions were expanded

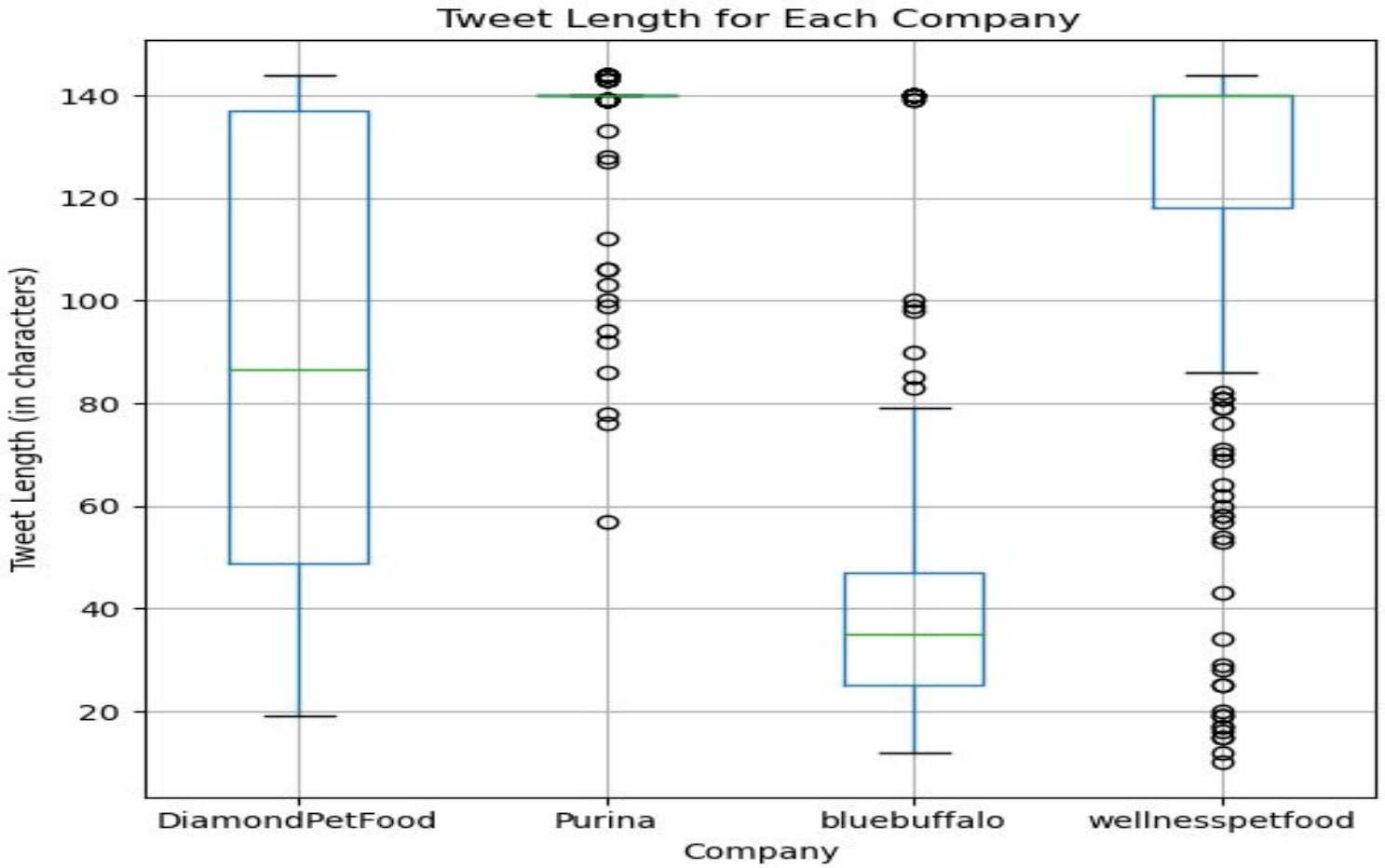
Cleaned DataFrame

	Tweet_Length	Tweet_Date	UserName	Full_Tweets	Tweet_Links	at_word	Mention_Count
0	140	2023-02-09 19:42:43+00:00	Purina	thank you for reaching out we re sorry that ...	https://t.co/k5ncW6PGIU	lhartness	1
1	139	2023-02-09 19:41:43+00:00	Purina	thank you for reaching out we re sorry that ...	https://t.co/tABQff7epw	fireshadowed	1
2	140	2023-02-09 16:09:53+00:00	Purina	thank you for reaching out to us when you ha...	https://t.co/GMsyUvy5Oq	BonnieblueBlue	1

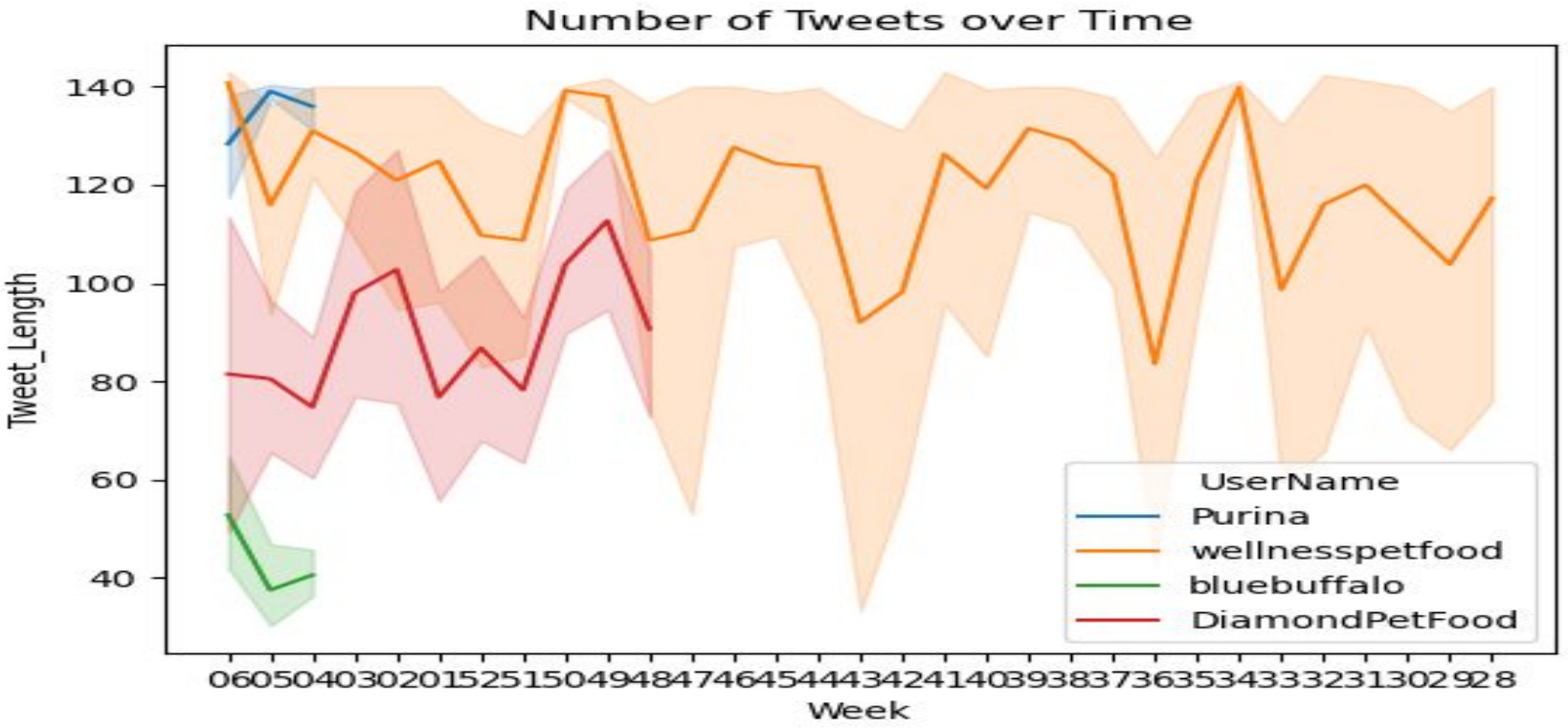
Exploratory Data Analysis

Tweet Length by Character

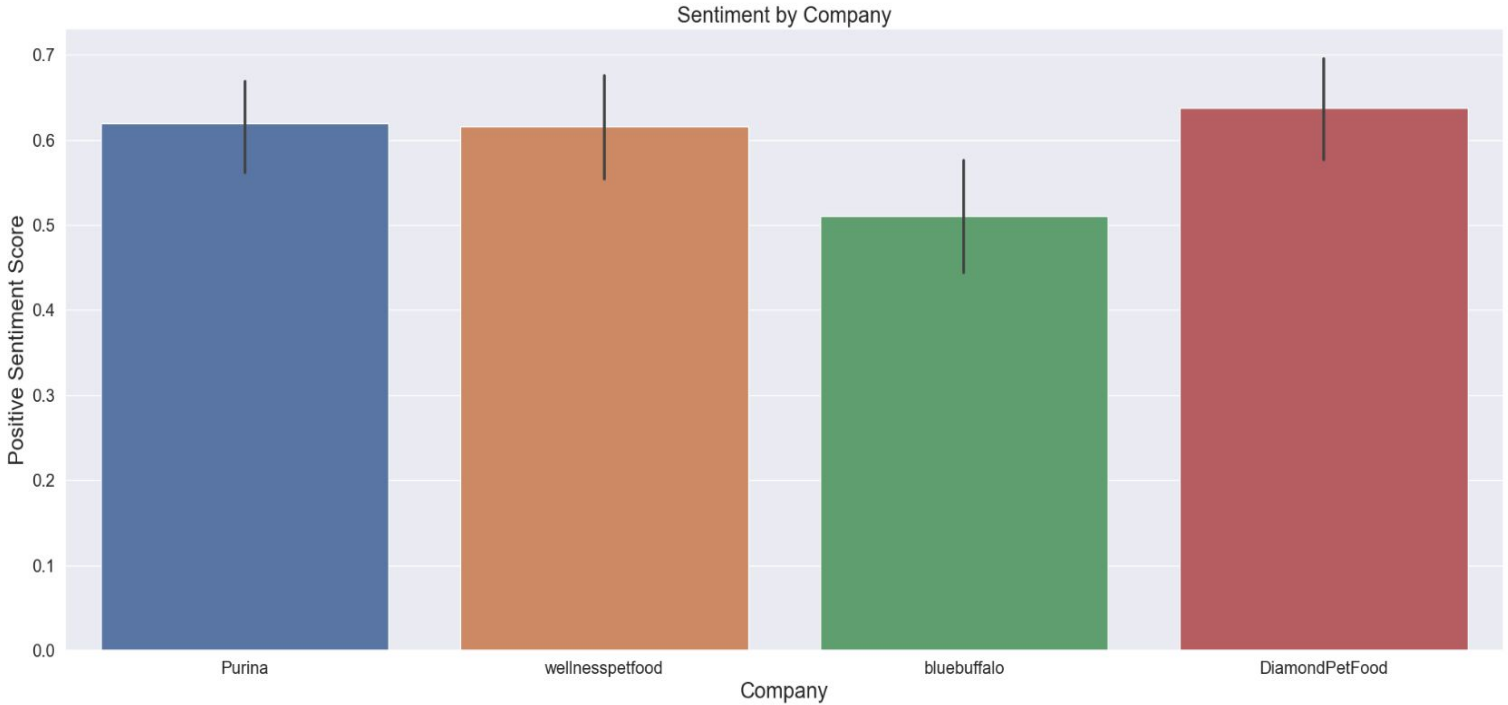



Tweet Length for Each Company
Tweet Length (in characters)
140
120
100
80
60
40
20
DiamondPetFood
Purina
bluebuffalo
wellnesspetfood
Company

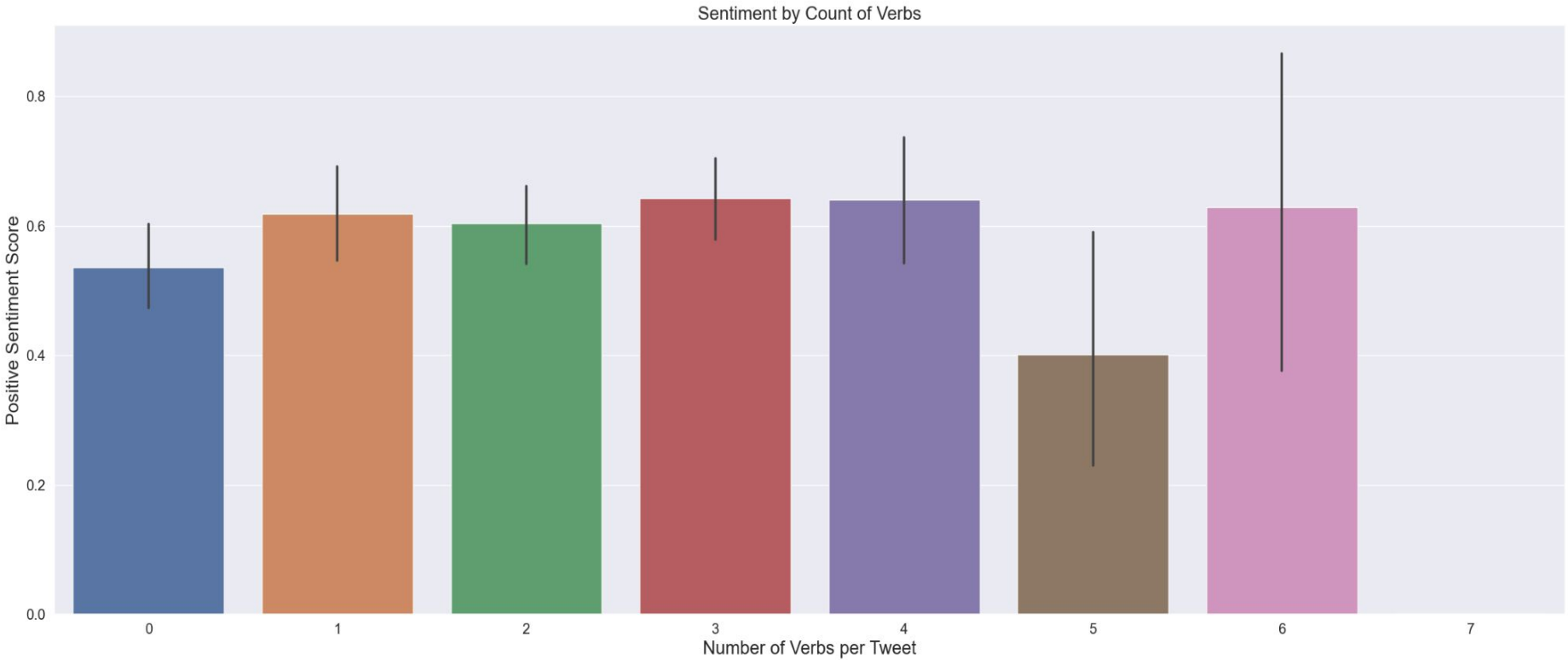
Tweet Length by Time to 200 Tweets



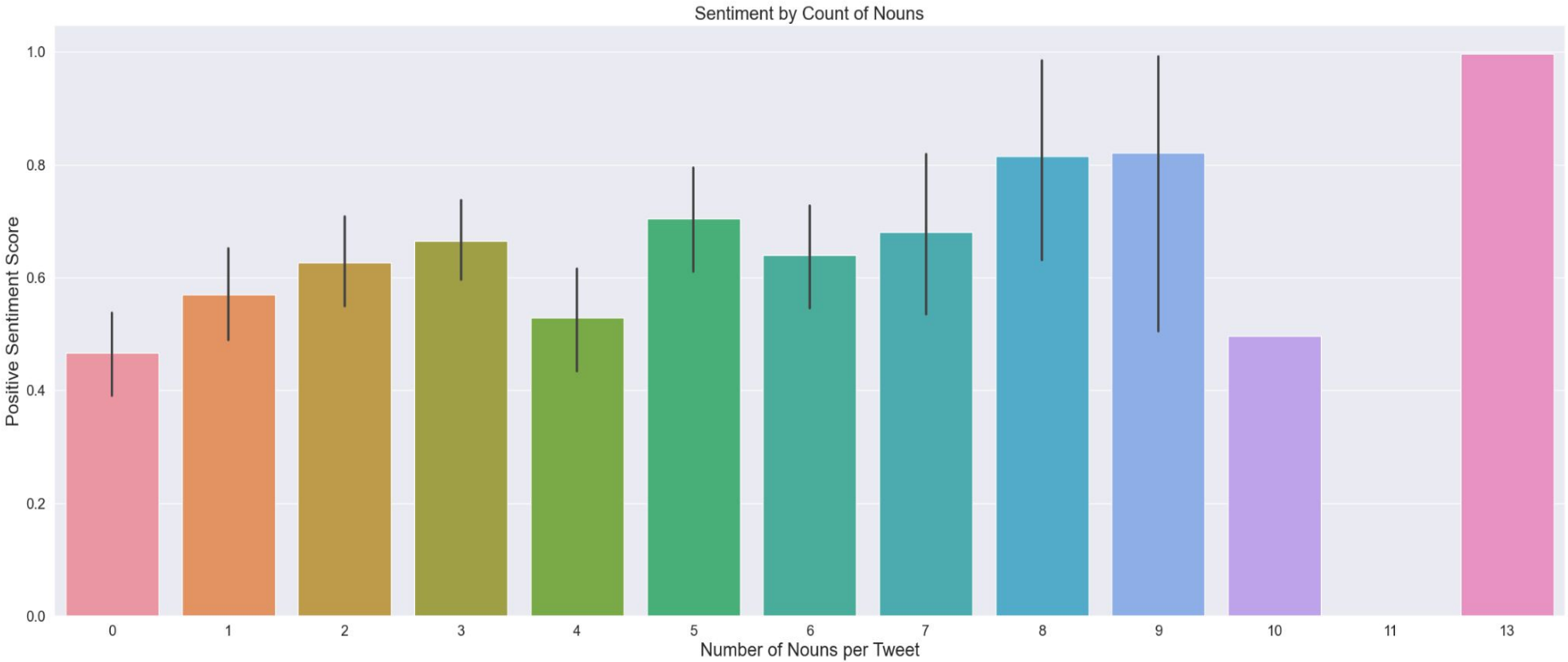
Positive Sentiment by Company



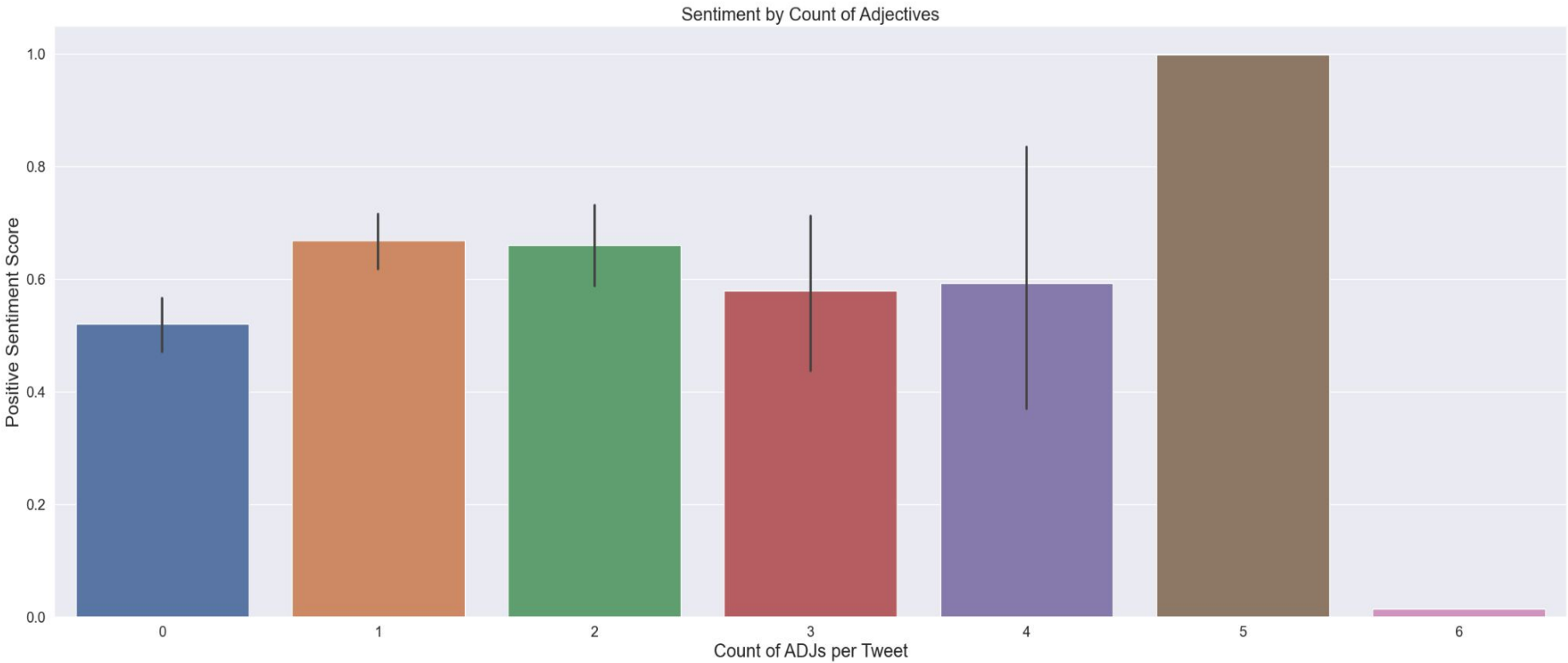
Positive Sentiment by Number of Verbs



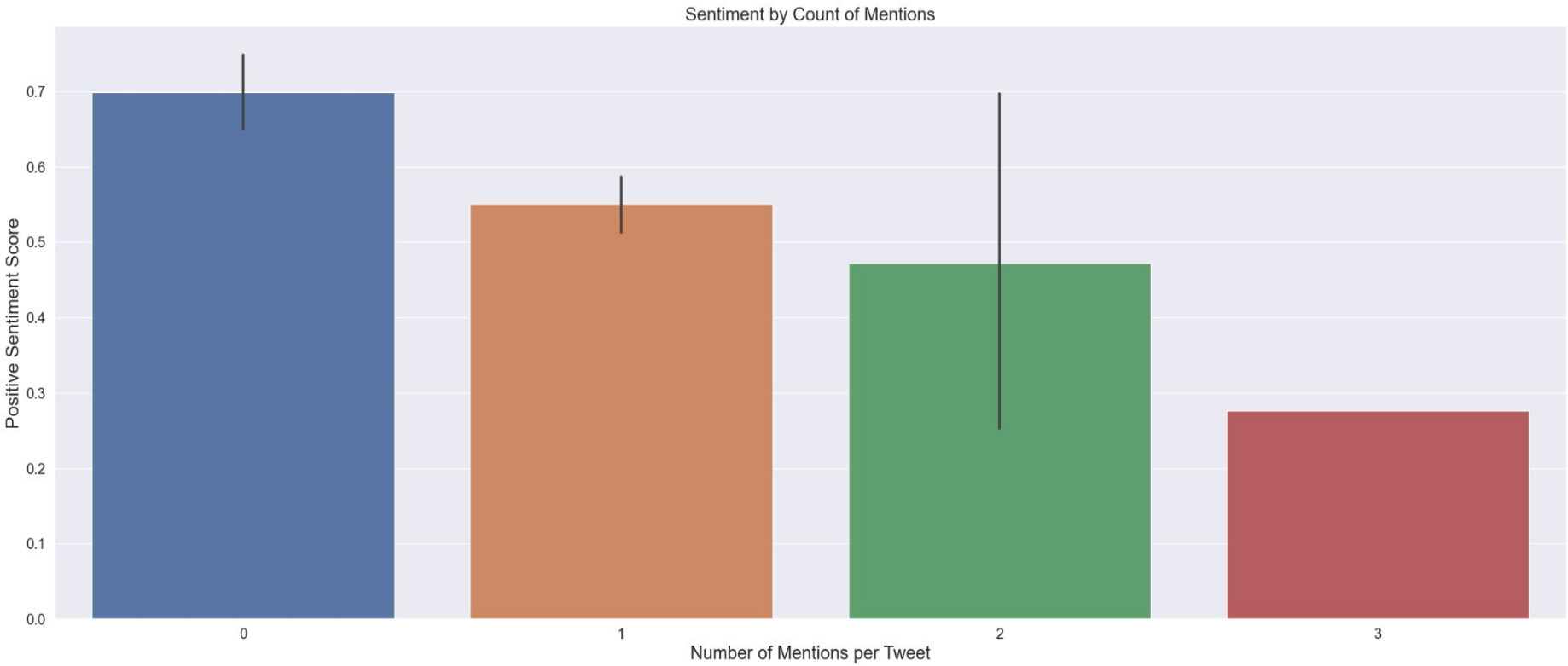
Positive Sentiment by Number of Nouns



Positive Sentiment by Number of Adjectives

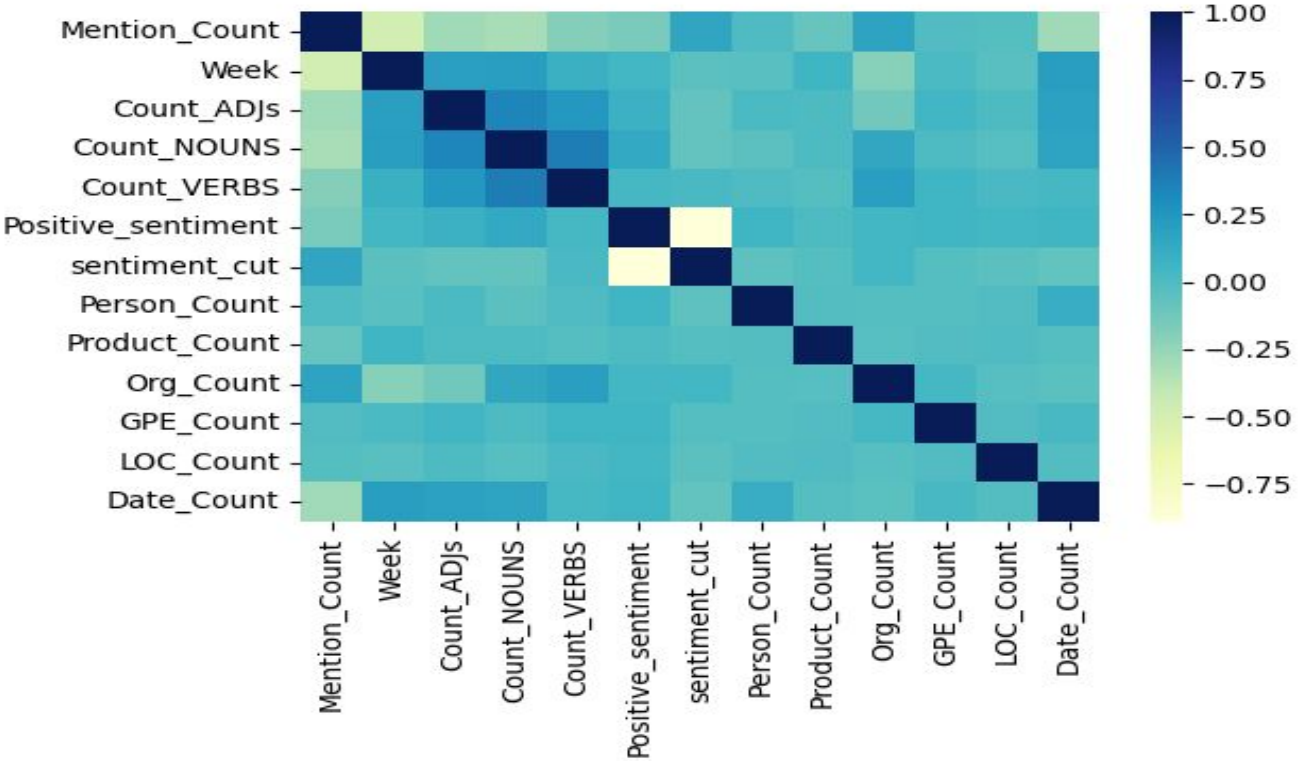


Positive Sentiment by Number of Mentions

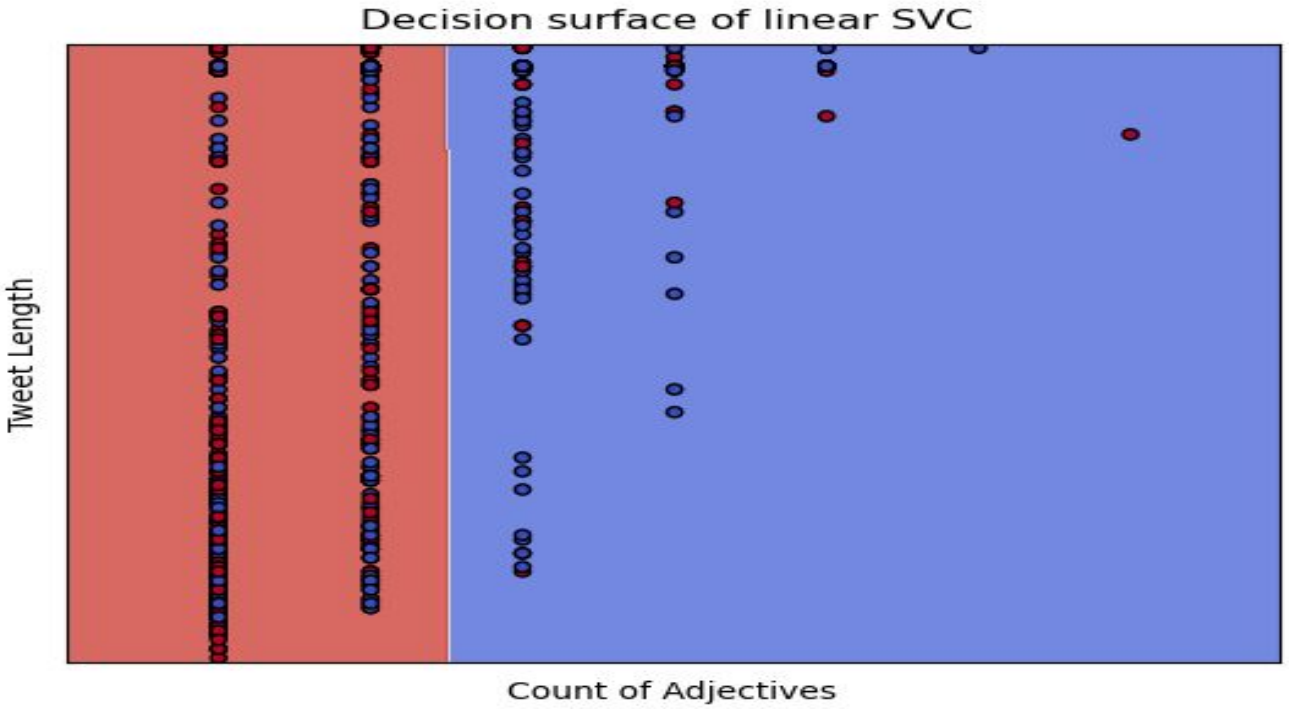


Predictive analysis with ML models

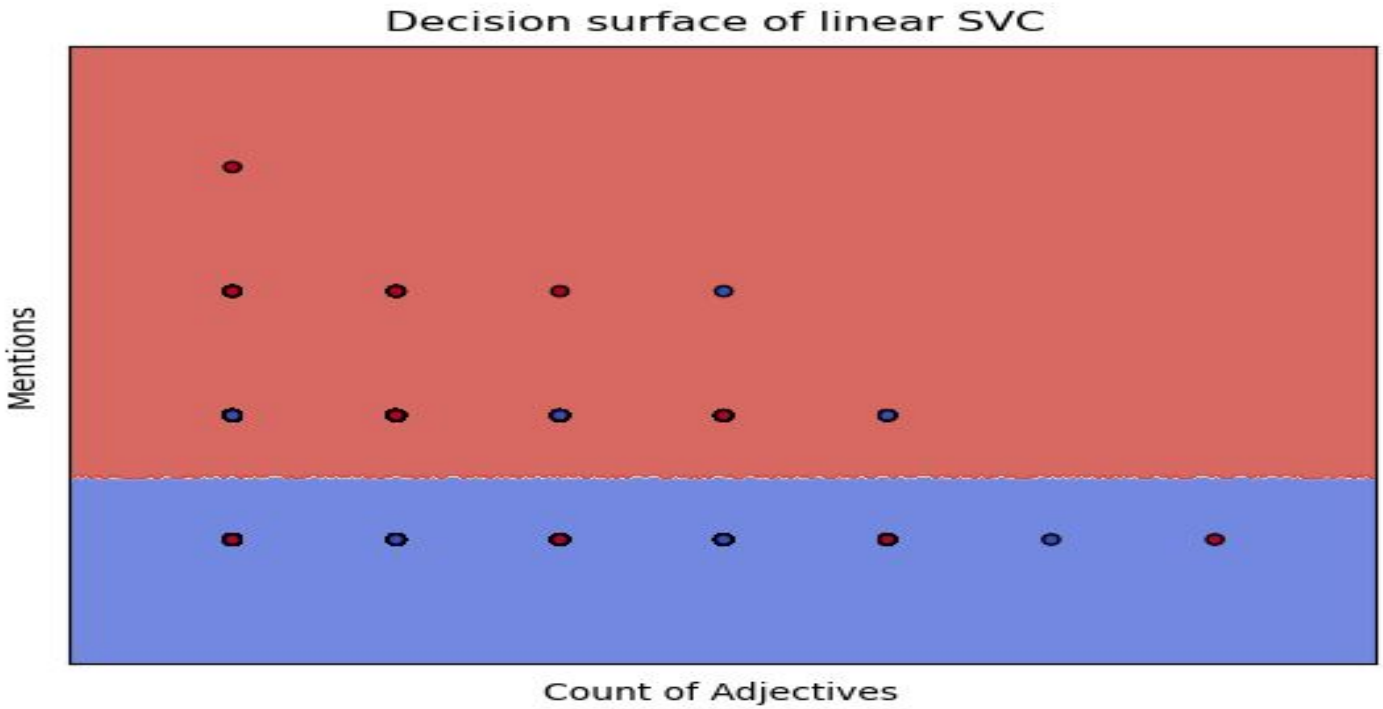
Correlation Heatmap



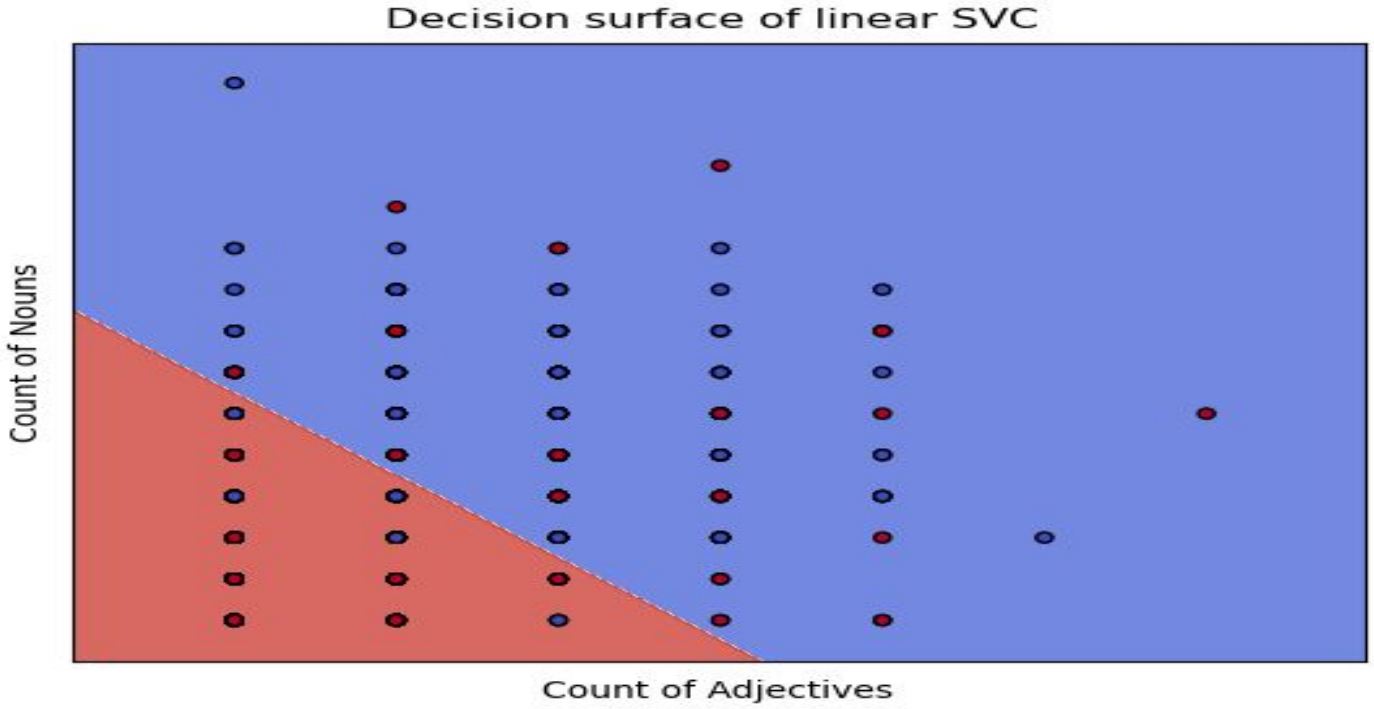
Linear SVC Tweet Length by Adjectives



Linear SVC Mentions by Adjectives



Linear SVC Nouns by Adjectives



Feature Engineering and the Better Model

Features in Final Model

Target variable: sentiment_cut (1 or 0 based on rounding of Positive sentiment score)

- UserName
- Mention_Count
- Count_ADJs
- Count_NOUNS
- Count_VERBS

Attempted Features with Negative Effects

- Week (converted as sine, cosine)
- Person_Count
- Product_Count
- Org_Count
- GPE_Count
- LOC_Count
- Date_Count
- Vocab_Size
- Diversity (type/token)

The (not too impressive) Model Metrics for Pet Food Sentiment Analysis

Model	Mean Absolute Error	Root Mean Squared Error	Testing Score for Positive Sentiment
Linear SVC	0.39	0.63	.61
Decision Tree Regressor max depth = 6	0.42	0.47	.10
Logistic Regression	0.39	0.63	.61
KNeighborsClassifier	0.41	0.64	.59
XGBClassifier	0.37	0.61	.63
XGB	0.46	0.48	.65